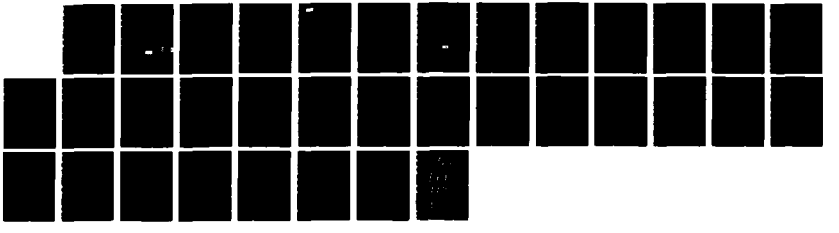
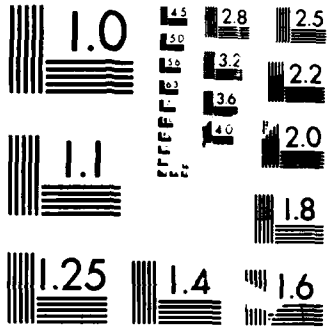
AD-A187 519 EQUIVALENT-GROUPS VERSUS SINGLE-GROUP EQUATING DESIGNS 1/1
FOR THE ACCELERATE (U) CENTER FOR NAVAL ANALYSES
ALEXANDRIA VA MARINE CORPS OPERATIO P H STOLOFF
UNCLASSIFIED JAN 87 CRM-87-6 N00014-87-C-0001 F/G 5/6 NL





CRM 87-6 / January 1987

RESEARCH MEMORANDUM

EQUIVALENT-GROUPS VERSUS SINGLE-GROUP EQUATING DESIGNS FOR THE ACCELERATED CAT-ASVAB PROJECT

Peter H. Stoloff

A Division of CNA Hudson Institute

CENTER FOR NAVAL ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Work conducted under contract N00014-87-C-0001.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE A187579

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CRM 87-6	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Center for Naval Analyses 6b. OFFICE SYMBOL (*If applicable*) CNA	7a. NAME OF MONITORING ORGANIZATION Commandant of the Marine Corps (Code RDS)
6c. ADDRESS (*City, State, and ZIP Code*) 4401 Ford Avenue Alexandria, Virginia 22302-0268	7b. ADDRESS (*City, State, and ZIP Code*) Headquarters, Marine Corps Washington, D.C. 20380
8a. NAME OF FUNDING / ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (*If applicable*) ONR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-87-C-0001
8c. ADDRESS (*City, State, and ZIP Code*) 800 North Quincy Street Arlington, Virginia 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: 65153M; PROJECT NO: C0031; TASK NO: ; WORK UNIT ACCESSION NO:

11. TITLE (*Include Security Classification*)

Equivalent-Groups Versus Single-Group Equating Designs For The Accelerated CAT-ASVAB Project

12. PERSONAL AUTHOR(S)
Peter H. Stoloff

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Final	FROM TO	January 1987	34

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUB-GROUP	ACAP (Accelerated CAT-ASVAB Program), Aptitude tests, ASVAB (Armed Services Vocational Aptitude Battery), CAT (Computerized Adaptive Test), Comparison, Intelligence test, PP (Paper and Pencil Tests), Test methods, Test scores
05	10		

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

Computerized adaptive test (CAT) and paper-and-pencil (PP) ASVAB scores were compared and equated under equivalent-groups and counterbalanced, repeated-measures designs. The results were sensitive to the order of presentation of the two versions of the ASVAB. Different equatings emerged from the two designs.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Lt.Col. G. W. Russell	22b. TELEPHONE (*Include Area Code*) (202) 694-3491	22c. OFFICE SYMBOL RDS-40

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted

All other editions are obsolete

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

CNA CENTER FOR NAVAL ANALYSES
A Division of Hudson Institute 4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268 • (703) 824-2000

11 Jun 1987

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Memorandum 87-6

Encl: (1) CNA Research Memorandum 87-6, "Equivalent-Groups Versus Single-Group Equating Designs for the Accelerated CAT-ASVAB Project," by Peter H. Stoloff, Jan 1987

1. Enclosure (1) is forwarded as a matter of possible interest.

2. The Department of Defense may implement a computerized adaptive testing (CAT) version of the Armed Services Vocational Aptitude Battery (ASVAB). Before CAT can be implemented it must be equated to the pencil-and-paper version. This Research Memorandum reports analysis on the appropriateness of various equating designs.

William H. Sims
Director, Manpower and Training Program
Marine Corps Operations Analysis Group

Distribution List:
Reverse Page

Subj: Center for Naval Analyses Research Memorandum 87-6

Distribution List

SNDL

A6 HQMC RD&S (2 copies)

A6 HQMC MPR
Attn: Director, Manpower Plans and Policy Division (2 copies)

Other

Joint Service CAT-ASVAB Working Group (15 copies)

Defense Advisory Committee on Military Personnel Testing (8 copies)

CRM 87-6 / January 1987

EQUIVALENT-GROUPS VERSUS SINGLE-GROUP EQUATING DESIGNS FOR THE ACCELERATED CAT-ASVAB PROJECT

Peter H. Stoloff

Marine Corps Operations Analysis Group

A Division of CNA *Hudson Institute*

CENTER·FOR·NAVAL·ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

ABSTRACT

Computerized adaptive test (CAT) and paper-and-pencil (PP) ASVAB scores were compared and equated under equivalent-groups and counterbalanced, repeated-measures designs. The results were sensitive to the order of presentation of the two versions of the ASVAB. Different equatings emerged from the two designs.

EXECUTIVE SUMMARY

The introduction of a computerized test adaptive (CAT) version of the Armed Services Vocational Aptitude Battery (ASVAB) will require that the scores from the new test be equated to those from the paper-and-pencil (PP) forms. Two equating designs, or approaches, were being considered for collecting the scores during the Accelerated CAT-ASVAB Project (ACAP). One approach, the "single group" design, is to administer both the PP and CAT versions of the test to each examinee in random order. An alternative is the "equivalent groups" design, in which each examinee takes only one form of the test. Although the latter approach usually requires twice the sample size, it eliminates the possibility of order-of-administration effects that are inherent in the single-group design. The purpose of this analysis is to determine whether order effects which might bias the equating of PP and CAT scores are indeed present.

Using data collected with a counterbalanced single-group design during an early experimental implementation of CAT, two equivalent groups of examinees were formed. The groups consisted of recruits who were administered the tests in the following orders: PP-first, CAT-second; and CAT-first, PP-second.

ORDER EFFECTS

There was general evidence of an order-of-administration effect. Those taking CAT before PP tended to score lower on the subsequent PP test. Those taking PP before CAT tended to improve on the subsequent CAT test. The results are summarized in table I.

EQUATING

The equipercentile method was used to equate CAT to PP ASVAB under the two equating designs. The CAT test scores from those who took CAT first were equated to the PP scores of those taking PP first in order to represent the results expected from an equivalent-groups design. The test scores from the entire sample (both groups combined) were used to generate the equating to be expected from a single-group design. Equivalency of groups was ascertained by comparing the individuals in each group on an independent measure (AFQT measured earlier, at time of application).

TABLE I

DIFFERENCES IN ASVAB SUBTEST MEAN SCORES BETWEEN ORDER-OF-PRESENTATION GROUPS

	Difference in mean scores	
Subtest[a]	PP1 – PP2[b]	CAT1 – CAT2[c]
GS	– .010	0.085
AR	0.618*	0.375
WK	0.417*	– 0.350
PC	1.121*	– 0.104
NO	0.425	– 2.442*
CS	– 1.320*	– 2.443*
AS/AI	0.856*	– 0.108
MK	0.275	0.397
MC	0.929*	– 1.057*
EI	0.831*	– 0.497
AS/SI	0.848*	– 0.325

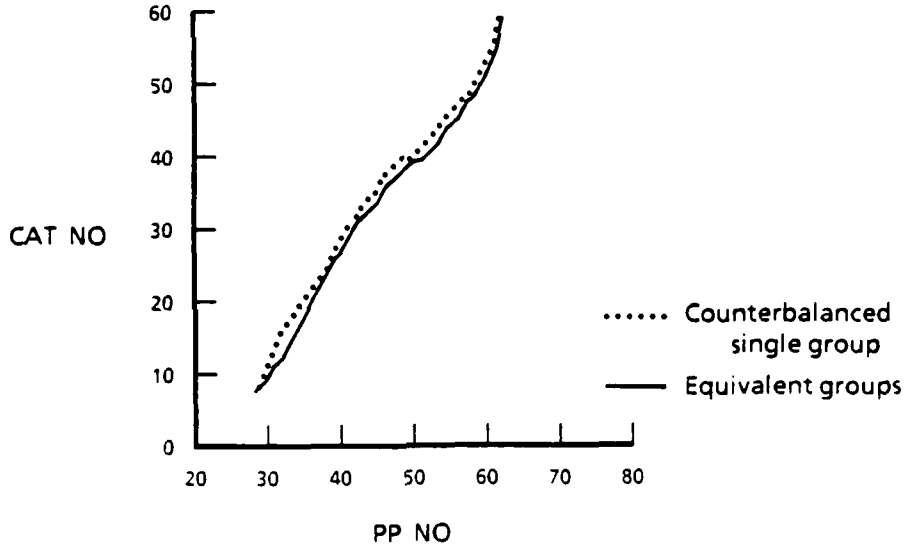
a. GS = General Science
AR = Arithmetic Knowledge
WK = Word Knowledge
PC = Paragraph Comprehension
NO = Numerical Operations
CS = Coding Speed
AS/AI = Auto Shop (PP), Automotive Information (CAT)
MK = Math Knowledge
MC = Mechanical Comprehension
EI = Electronics Information
AS/SI = Auto Shop (PP), Shop Information (CAT)

b. Average PP score for those taking PP before CAT, minus average PP score for those taking PP after CAT (ASVAB standard score metric).

c. Average CAT score for those taking CAT before PP, minus average CAT score for those taking CAT after PP. Power test scores shown in theta metric, standardized to mean = 50, standard deviation = 10, with orders combined. Speeded test scores are number correct, per unit time.

* Significantly different from zero at the .05 confidence level (t test).

Small but consistent differences, on the order of one or two ASVAB standard score points, as a function of type of equating design were observed. Counterbalancing tends to result in a higher CAT score for a given PP score, with some reversals at the ends of the scale. Figure I shows the equatings for the Numerical Operations subtest.



NOTE: CAT NO was scored as number correct per unit of time.

FIG. I: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: NUMERICAL OPERATIONS (NO) SUBTEST

IMPLICATIONS

The results suggest that the equivalent groups equating design is less problematic than the counterbalanced design and should be the one chosen for implementation during ACAP. When CAT becomes operational it will be administered by itself, as it would be in an equivalent-groups design. Conversion tables used to convert CAT scores to the 1980 ASVAB metric must be appropriate for this method of presentation. This will require an equivalent-groups design.

Higher CAT scores following practice on PP suggest the importance of proper "warm up" before taking CAT.

Taking the entire CAT battery may be fatiguing. Making the CAT battery longer, by seeding items in existing subtests or by including new experimental tests in the battery, should be viewed with caution.

TABLE OF CONTENTS

	Page
List of Illustrations	ix
List of Tables	xi
Background	1
Equating Designs	1
Purpose	2
Approach	2
Data	2
Test Administration	2
Test for Equivalency of Groups	3
CAT Item Pool	3
Speeded Test Scoring	3
Analytic Procedures	5
Equating Designs	5
Equating Method	5
Estimating Order-of-Presentation Effects	5
Results	6
Order Effects	6
Equating	7
Implications	8
References	15
Appendix: Equating Tables	A-1 – A-3

LIST OF ILLUSTRATIONS

		Page
1	Equipercentile Equating by Type of Design: General Science (GS) Subtest	9
2	Equipercentile Fquating by Type of Design: Arithmetic Reasoning (AR) Subtest	9
3	Equipercentile Equating by Type of Design: Word Knowledge (WK) Subtest	10
4	Equipercentile Equating by Type of Design: Paragraph Comprehension (PC) Subtest	10
5	Equipercentile Equating by Type of Design: Numerical Operations (NO) Subtest	11
6	Equipercentile Equating by Type of Design: Coding Speed (CS) Subtest	11
7	Equipercentile Equating by Type of Design: Automotive Information (AI) Subtest	12
8	Equipercentile Equating by Type of Design: Math Knowledge (MK) Subtest	12
9	Equipercentile Equating by Type of Design: Mechanical Comprehension (MC) Subtest	13
10	Equipercentile Equating by Type of Design: Electronics Information (EI) Subtest	13
11	Equipercentile Equating by Type of Design: Shop Information (SI) Subtest	14

LIST OF TABLES

		Page
1	Mean MEPS Scores by Order of Subsequent Test Administration	4
2	Equivalent-Groups Testing Paradigm	6
3	Differences Between Group Subtest Mean Scores	7

BACKGROUND

The introduction of a computerized adaptive test (CAT) version of the Armed Services Vocational Aptitude Battery (ASVAB) will require that the scores from the new test be equivalent to those from the paper-and-pencil (PP) forms. The process of making scores from different forms of the same test equivalent is known as equating. An equating method that has frequently been used for ASVAB is "equipercentile equating." This method will also be used in the Accelerated CAT-ASVAB Program (ACAP), when CAT ASVAB will be equated to PP ASVAB in an operational environment. In order to perform the equating, distributions of scores for both tests must be constructed. Two equating designs, or approaches, were being considered for collecting the scores.

Equating Designs

One approach, the "single group" design, is to administer both the PP and CAT versions of the test to each examinee in random order. An advantage of this approach is that fewer examinees are required than with the alternative "equivalent groups" design, in which each examinee takes only one form of the test. The major disadvantage of the single-group design is that taking one test can affect the scores on the second test.

Of particular concern are asymmetrical order-of-presentation effects. This occurs when taking test "A" first affects the score on test "B" to a different degree than taking test "B" first affects the scores on "A." For example, suppose that taking the CAT before taking the PP test depresses the PP scores, while taking the PP test before the CAT has no effect on the CAT score. This phenomenon might arise from a flickering CRT, which could cause fatigue or eye strain and thereby affect subsequent test scores. Even with a counterbalanced design[1] PP scores collected after CAT would be lower than expected, and their inclusion in the PP score distribution would bias the equating.

There is reason to suppose that asymmetrical order-of-presentation effects are a real concern. A recent (unpublished) study by researchers at the University of Minnesota, reported by the Air Force Human Resources Laboratory (AFHRL), found an order effect for the Paragraph Comprehension (PC) ASVAB subtest with male Air Force

1. Equal numbers of examinees are randomly preassigned to each order of testing group.

recruits [1]. Those examinees who were tested first on a computer subsequently achieved lower scores when tested with a PP version of PC. No such effect was observed for those taking the PP test before being tested with the computer.

Purpose

The purpose of this research is to examine the likely effect of the equating design on the equating process that must be completed before a CAT version of ASVAB is implemented. The determination of which equating design to use is based on a comparison of equatings resulting from both single-group and equivalent-groups designs. If identical equatings are produced with the two designs, then either could be used during ACAP without fear of biased results. If differences attributable to order effects are observed, then a choice must be made.

Approach

A properly constructed counterbalanced single-group equating design will also incorporate an equivalent-groups design. This occurs because half of the examinees tested with a single-group design will have taken CAT before PP, and the other half will have taken PP first. Thus, the CAT scores of one group should be unaffected by the PP test, and the PP scores of the other group should be unaffected by having taken CAT. If examinees have been randomly assigned a testing order, then the sample should contain two equivalent groups.

DATA

Test Administration

The data used in this analysis were collected by the Navy Personnel Research and Development Center (NPRDC) as part of a joint armed services study to validate an early version of CAT [2]. In that study, the CAT scores were validated against training grades. Both CAT and PP ASVAB retest data (ASVAB forms 8, 9, or 10) were collected during initial skill training from about 7,000 recruits representing all of the armed services. The PP and CAT tests were administered back-to-back in counterbalanced order with a short intervening period. The CAT data consisted of a complete CAT battery (that is, all subtests), whereas the PP data included only those PP subtests

required for selection into the particular military occupational specialty (MOS) of the recruit. PP subtests typically numbered from three to five per individual.

Test for Equivalency of Groups

The retest data were supplemented by test scores collected earlier at the time of application to the armed forces at Military Entrance Processing Stations (MEPS). The MEPS scores were used as a check for equivalent groups. (Equivalent groups should have comparable MEPS scores for a given ASVAB subtest.)

Although the data were not collected explicitly to produce equivalent groups, it appears that the groups were in fact equivalent.

Two tests of statistical significance of differences between MEPS scores of those in the different order-of-presentation groups were performed (table 1). The Student's *t* statistic was used to compare the mean scores and the Kolmogorov-Smirnov (K-S) two-sample statistic used to compare the distributions of scores between order groups. The results indicate that the distributions of Math Knowledge (MK) scores for the two groups were somewhat different; otherwise the groups appear to be equivalent with respect to the other MEPS subtest score distributions. There were no statistically significant differences between mean MEPS scores. These results are taken as evidence of equivalence of the two order-of-presentation groups—that is, the groups should have the same underlying ability distribution.

CAT Item Pool

The CAT item pool consisted of about 200 items per subtest. The items were calibrated by NPRDC on a sample of about 2,500 applicants using LOGIST II. The CAT items were both selected and scored using Owen's Bayes procedure [3].

Speeded Test Scoring

The speeded tests, Numerical Operations (NO) and Coding Speed (CS), were administered by computer but conventionally scored as number correct during the original data collection. These subtests were subsequently rescored as number correct

per unit time; that is, 4 times number correct divided by testing time (minutes).[1] This scoring method is consistent with that expected to be used in the ACAP; it tends to increase the variability at the high end of the score continuim because some examinees finish the test before the time limit. The "rate" scores were used in the equating.

TABLE 1

MEAN MEPS SCORES BY ORDER OF SUBSEQUENT TEST ADMINISTRATION

	Order group			
Subtest[a]	PP1[b]	PP2[c]	Difference	χ^2 [d]
GS	56.38	56.60	−0.22	0.59
AR	55.22	55.43	−0.21	2.47
WK	51.47	51.32	0.15	1.89
PC	52.32	52.17	0.15	1.48
NO	54.25	53.80	0.45	0.68
CS	53.01	53.15	−0.14	1.57
AS	57.29	57.33	−0.04	0.94
MK	56.13	56.16	−0.03	7.20*
MC	55.00	54.87	0.13	0.87
EI	56.45	56.19	0.26	0.94

a. GS = General Science
AR = Arithmetic Knowledge
WK = Word Knowledge
PC = Paragraph Comprehension
NO = Numerical Operations
CS = Coding Speed
AI = Automotive Information
MK = Math Knowledge
MC = Mechanical Comprehension
EI = Electronics Information
SI = Shop Information

b. Those who later took PP ASVAB before CAT.

c. Those who later took PP ASVAB after CAT.

d. Chi-square approximation of K-S statistic (2 degrees of freedom).

* Significant at the 0.05 confidence level (K-S statistic).

1. A scalar of 4 was chosen to spread the resulting scores over an approximate range of 0-70.

ANALYTIC PROCEDURES

Equating Designs

The data used in this analysis came from a counterbalanced single-group administration of CAT and PP ASVAB. The general approach was to identify the order of administration of CAT and PP tests for each examinee and then form two equivalent groups on the basis of test order. A comparison of scores for those who took a particular test first (i.e., CAT or PP) with the scores of those taking that same test after the other test was used to identify order-of-administration effects. CAT was equated to PP for the sample as a whole and for the "CAT-first" and "PP-first" samples. This allowed a comparison of single-group versus equivalent-groups equating results.

Equating Method

Equipercentile equating, using analytic three-point moving-average smoothing of raw frequencies, was used to equate the CAT to the PP tests. CAT scores, although originally estimated in the "theta" metric, were linearly transformed to a scale with a mean of 50 and a standard deviation of 10. These transformed scores were rounded to the nearest integer. The transformation was necessary to form a discrete distribution, as required by the computer software used to perform the equating. The result is a scale with unit intervals corresponding to one-tenth of a standard deviation of the original theta metric.

Estimating Order-of-Presentation Effects

As defined earlier, order-of-presentation effects can be inferred when the scores for a given test depend on the order in which the test was administered. Using the available data, it is possible to compare PP test scores obtained when PP was given before CAT to those obtained when PP was given after CAT. Similar comparisons of scores on CAT presented before and following a PP test could be used to determine the effects of a PP test on a subsequent CAT test. Differences in mean test scores using the paradigm illustrated in table 2 are used as indicators of order effects. For example, the mean PP score for those in group 1 (PP taken first, or PP1) is compared with the mean PP score of those in group 2 (PP following CAT, or PP2). A positive value of the mean difference, PP1 – PP2, would indicate that taking CAT before PP lowers, or inhibits,

the subsequent PP score. A negative difference would indicate that taking CAT first facilitates the subsequent PP score.

TABLE 2

EQUIVALENT-GROUPS TESTING PARADIGM

	Order of administration	
Group	**First**	**Second**
1	PP	CAT
2	CAT	PP

RESULTS

Order Effects

Mean differences in equivalent-group PP and CAT subtest scores are shown in table 3. Statistically significant differences at the .05 confidence level are indicated by asterisks.

A comparison of the mean difference in retest scores (PP, as well as CAT) between groups show small but significant order effects. Scores on paper-and-pencil tests taken after CAT tend to be lower than PP-taken-first scores, suggesting that CAT may interfere with PP. An exception was the Coding Speed (CS) test. Paper-and-pencil CS was the only test positively affected when preceded by CAT. Test scores for both speeded CAT tests (NO and CS) tended to improve when preceded by the PP versions.

TABLE 3

DIFFERENCES BETWEEN GROUP SUBTEST MEAN SCORES

	Differences in mean scores	
Subtest	PP1 – PP2[a]	CAT1 – CAT2[b]
GS	–0.010	0.085
AR	0.618*	0.375
WK	0.417*	–0.350
PC	1.121*	–0.104
NO	0.425	–2.442*
CS	–1.320*	–2.443*
AS/AI[c]	0.856*	–0.108
MK	0.275	0.397
MC	0.929*	–1.057*
EI	0.831*	–0.497
AS/SI[c]	0.848*	–0.325

a. ASVAB standard score metric.
b. Theta metric, standardized to mean = 50, standard deviation = 10, with orders combined.
c. Auto Shop (PP) and Auto Information (CAT); Auto Shop (PP) and Shop Information (CAT).
* Significant at the .05 confidence level (t test).

Equating

Figures 1 through 11 show the equipercentile equating of the CAT to PP subtests in the battery resulting from the two equating designs—that is, equivalent-groups and counterbalanced single-group designs. The general pattern of results again indicate a small, but consistent, design effect. The single-group design results in higher CAT scores for a given PP score, with some reversals at the low end of the score scale. The most pronounced effects are again with the speeded tests—NO, in particular.

IMPLICATIONS

The results suggest that the particular design chosen for equating CAT to PP ASVAB can have a noticeable effect on the resulting conversion tables (see appendix). A counterbalanced single-group design was shown to be prone to order effects and resulted in higher CAT scores for a given PP score.

The observation that CAT scores can improve with prior practice, although on a different medium of presentation, suggests the importance of a proper "warm-up" for CAT test takers.

The general pattern of results showed that taking PP after CAT resulted in lower PP scores, suggesting that taking a full CAT battery may be fatiguing. Further lengthening a CAT battery (for example, by seeding extra items for purposes of on-line calibration) could affect operational test scores.

When CAT becomes operational it will be administered by itself (that is, as it would be in an equivalent-groups design). The conversion tables used to convert the theta to the 1980 metric must be appropriate for this mode of presentation. This will require an equivalent-groups design.

Considering all of the results, an equivalent-groups design appears to be less problematic compared to a counterbalanced single-group design.

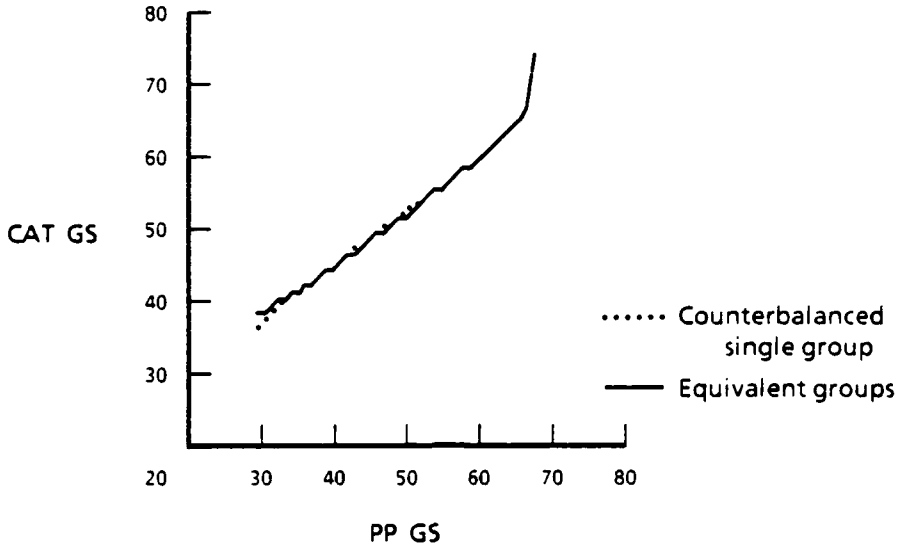


FIG. 1: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: GENERAL SCIENCE (GS) SUBTEST

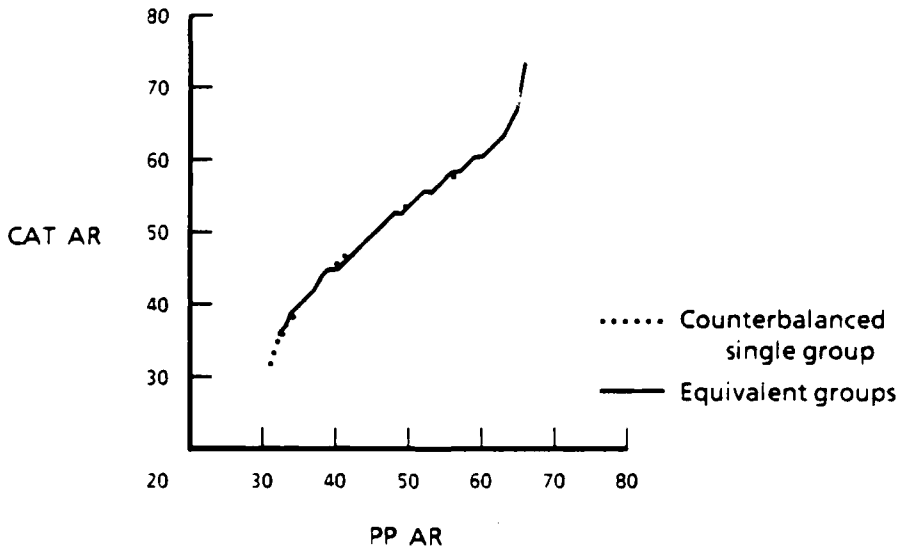


FIG. 2: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: ARITHMETIC REASONING (AR) SUBTEST

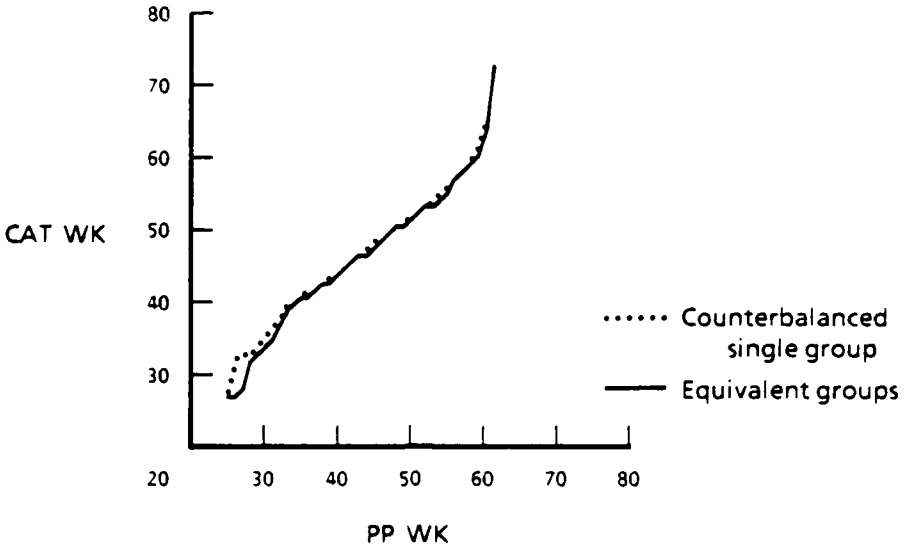


FIG. 3: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: WORD KNOWLEDGE (WK) SUBTEST

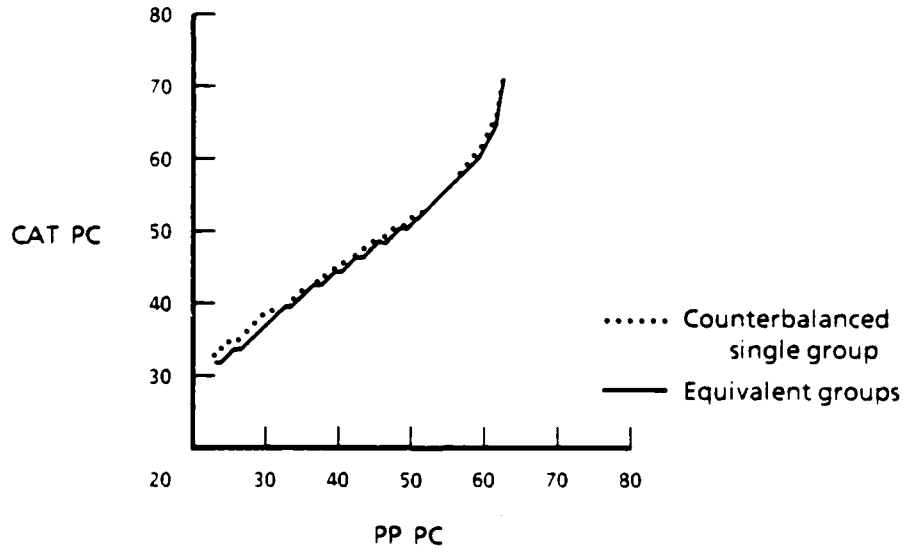


FIG. 4: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: PARAGRAPH COMPREHENSION (PC) SUBTEST

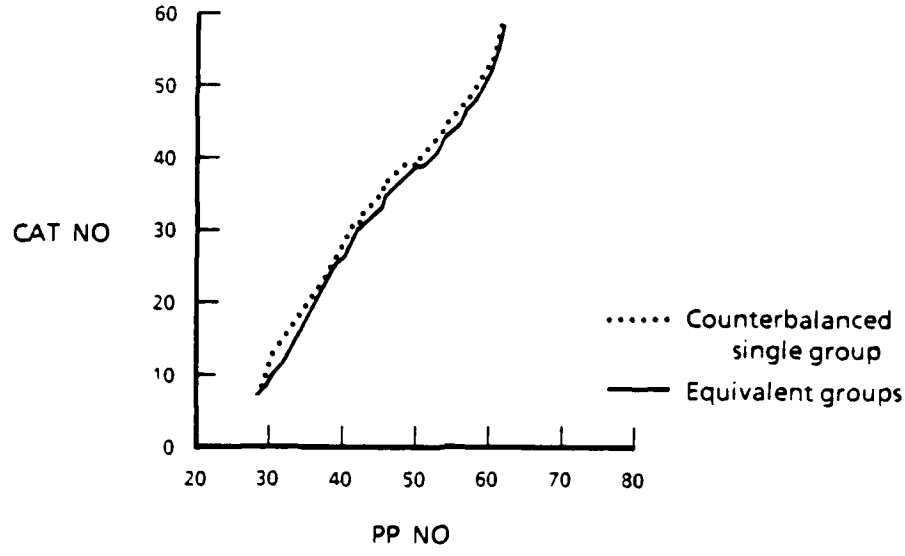


FIG. 5: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: NUMERICAL OPERATIONS (NO) SUBTEST

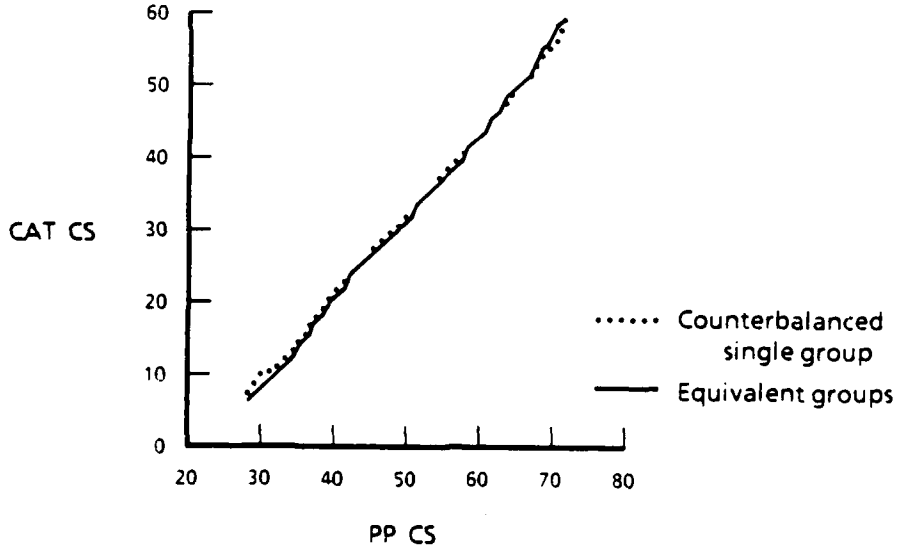


FIG. 6: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: CODING SPEED (CS) SUBTEST

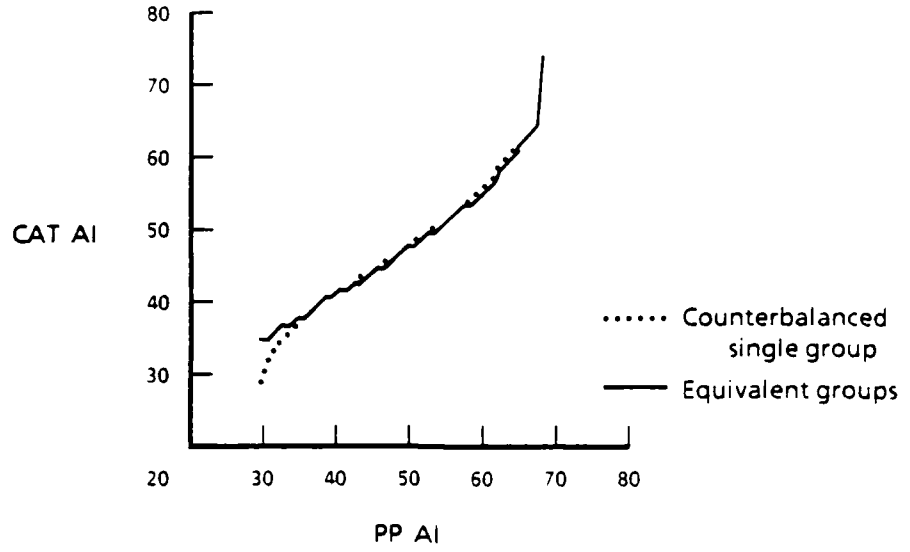


FIG. 7: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: AUTOMOTIVE INFORMATION (AI) SUBTEST

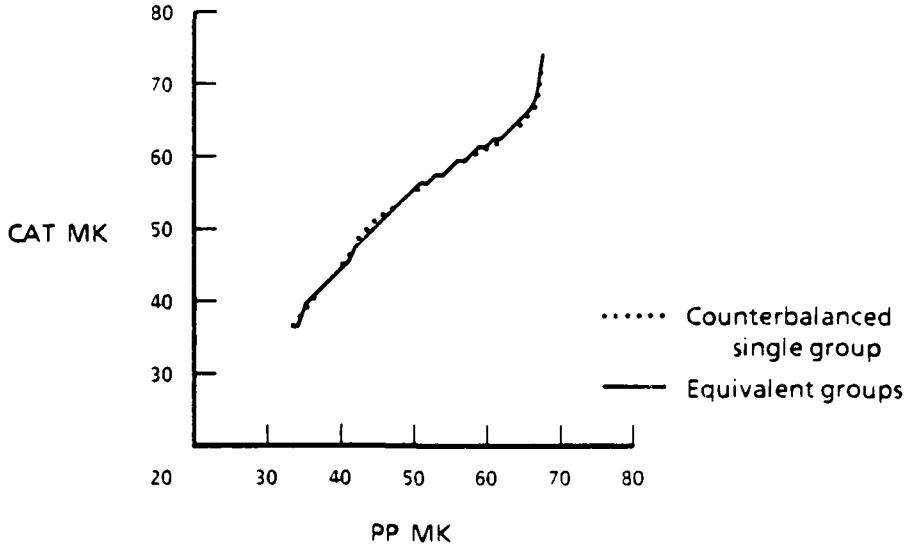


FIG. 8: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: MATH KNOWLEDGE (MK) SUBTEST

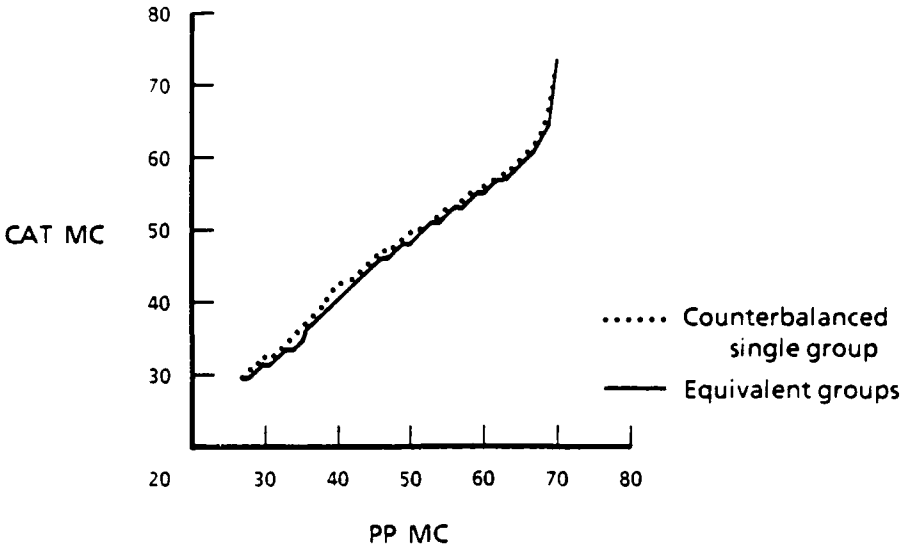


FIG. 9: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: MECHANICAL COMPREHENSION (MC) SUBTEST

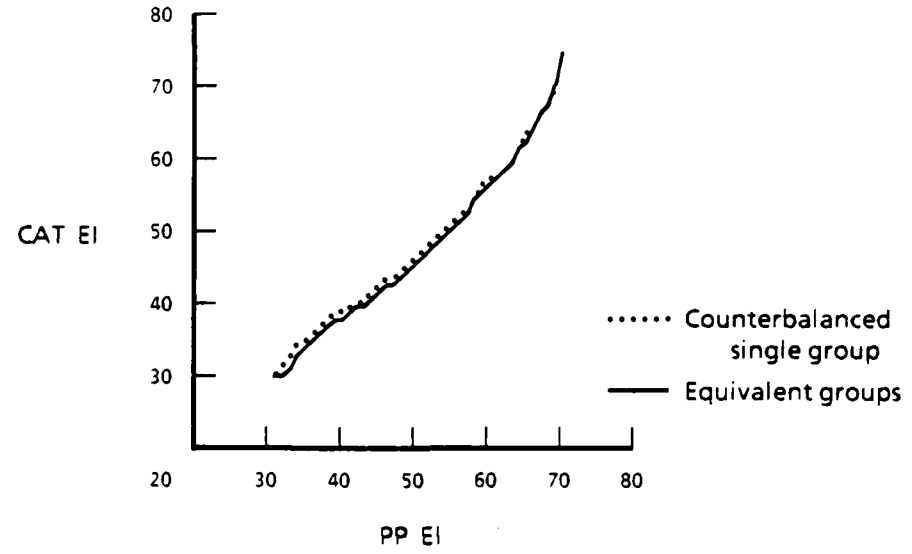


FIG. 10: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: ELECTRONICS INFORMATION (EI) SUBTEST

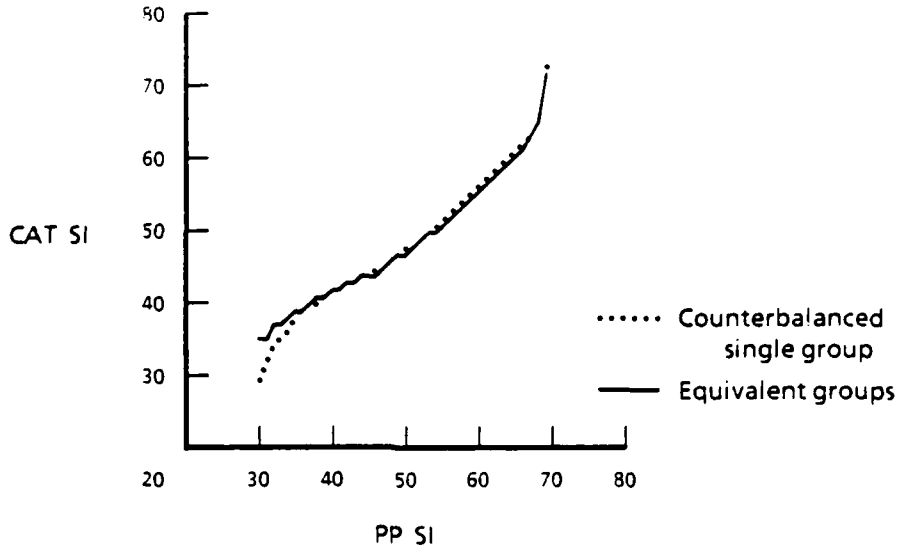


FIG. 11: EQUIPERCENTILE EQUATING BY TYPE OF DESIGN: SHOP INFORMATION (SI) SUBTEST

REFERENCES

[1] Air Force Human Resources Laboratory, "Service Data Bearing on ACAP Equating," by James Earles, as contained in CNA 86-0446, "Minutes of December 1985 Meeting of CAT-ASVAB Psychometric Committee," by William Sims, Mar 1986

[2] Rehab Group, Inc., "Predictive Validity Evaluation of Computerized Adaptive Testing: Results of the Navy Research," by Susan B. Hardwick and Kenneth D. White, Dec 1983

[3] Roger J. Owen. "A Bayesian Sequential Procedure for Quantal Response in the Context of Adaptive Mental Testing." *Journal of the American Statistical Association* (Jun 1975): 351-356

APPENDIX

EQUATING TABLES

APPENDIX

EQUATING TABLES

CAT subtest scores equivalent to PP ASVAB standard subtest scores are shown in table A-1 for the two equating designs. Scores are in the 1980 metric. The CAT scores were linearly transformed, using an additive constant of 50 and a multiplier of 10. This transformation results in a range of scale values of 20 to 80, which corresponds to a range of ±3 in the original theta metric.

Sample sizes used in the equating designs are shown in table A-2.

TABLE A-1

EQUIVALENT CAT SCORES BY EQUATING DESIGN

CAT Score

PP Score	GS		AR		WK		PC		NO		CS		AS		MK		MC		EI		SI	
	SG	EG	SG	EG	SG	EG	SG	EG	SG	EG	SG	EG	SG	EG	SG	EG	SG	EG	SG	EG	SG	EG
20																						
21							32															
22							33			4												
23							34	32		6												
24							35	33		7												
25					28		35	34		7												
26					32	27	36	34	4	7												
27					33	28	37	35	6	7							29					
28					33	32	38	36	8	8	7	6					30	29				
29	37				34	33	39	37	12	9	9	7					31	30	28			
30	38	39			35	34	39	38	14	11	10	8	29				32	31	29		29	
31	39	40	32		37	35	40	39	16	12	10	9	32	35			32	31	30		32	35
32	40	41	35	36	38	37	40	40	17	14	11	10	34	36			33	32	32	30	34	37
33	41	41	37	37	40	39	41	40	19	16	12	11	35	37	37		34	33	33	31	35	37
34	41	42	38	39	40	40	42	41	20	18	13	12	36	37	38	37	35	33	35	33	36	38
35	42	42	40	40	41	41	42	42	22	20	15	14	37	38	39	40	36	34	35	34	38	39
36	42	43	41	41	42	41	43	43	23	22	16	15	38	38	41	41	37	36	36	35	39	39
37	43	43	42	42	42	42	44	43	25	24	18	17	39	39	42	42	38	37	37	36	40	40
38	44	44	44	44	43	43	45	44	27	26	19	18	40	40	43	43	39	38	38	37	40	41
39	45	45	45	45	44	43	45	45	29	27	21	20	41	41	44	44	41	39	39	38	41	41
40	45	45	46	45	44	44	46	45	31	29	22	21	41	41	45	45	42	40	39	38	42	42
41	46	46	47	46	45	45	47	46	32	31	23	22	42	42	47	46	43	41	40	39	42	42
42	47	47	47	47	46	46	47	47	34	32	24	24	42	42	48	48	43	42	40	40	43	43
43	48	47	48	48	47	47	48	47	35	33	25	25	43	43	50	49	44	43	41	40	43	43
44	48	48	49	49	48	47	49	48	36	34	26	26	44	43	51	50	45	44	42	41	44	44
45	49	49	50	50	49	48	49	49	38	36	28	27	44	44	52	51	46	45	43	42	44	44
46	50	50	51	51	49	49	50	49	39	37	29	28	45	45	53	52	47	46	44	43	45	44
47	51	50	52	52	50	50	51	50	40	38	30	29	46	45	53	53	47	46	44	43	45	45
48	51	51	53	53	51	51	51	51	40	39	31	30	46	46	54	54	48	47	45	44	46	46
49	52	52	54	53	52	51	52	51	41	40	32	31	47	47	55	55	49	48	46	45	47	47
50	53	52	54	54	52	52	53	52	42	40	33	32	48	48	56	56	49	48	47	46	48	47
51	54	53	55	55	53	53	53	53	43	41	34	34	49	48	56	57	50	49	48	47	48	48
52	54	54	56	56	54	54	54	54	44	42	35	35	49	49	57	57	51	50	49	48	49	49
53	55	55	56	56	55	54	55	55	46	44	36	36	50	50	58	58	51	51	50	49	50	50
54	56	56	57	57	56	55	56	56	47	45	38	37	51	50	58	58	52	51	51	50	51	50
55	56	56	58	58	57	56	57	57	48	46	39	38	51	51	59	59	53	52	52	51	52	51
56	57	57	58	59	58	58	58	58	49	48	40	39	52	52	60	60	53	53	53	52	53	52
57	58	58	59	59	59	59	60	59	51	49	41	40	53	53	60	60	54	53	54	53	54	53
58	59	59	60	60	60	60	61	60	53	51	42	42	54	54	61	61	55	54	55	55	55	54
59	59	59	61	61	62	61	62	61	55	53	43	43	55	54	61	62	55	55	57	56	56	55
60	60	60	61	61	65	64	64	63	58	56	44	44	56	55	62	62	56	55	58	57	57	56
61	61	61	62	62	74	74	66	65	62	61	46	46	57	56	62	63	57	56	58	58	58	57
62	62	62	63	63		72	72				47	47	58	57	63	63	57	57	59	59	59	58
63	63	63	64	64							48	49	60	59	64	64	58	57	60	60	60	59
64	64	64	66	66							50	50	61	60	65	65	59	58	62	62	61	60
65	65	65	68	68							51	51	62	61	65	66	60	59	64	63	62	61
66	66	66	74	74							52	52	63	63	67	67	61	60	65	65	63	62
67	68	68									53	54	64	64	68	69	62	61	67	67	64	64
68	75	75									55	56	66	65	75	75	64	63	68	68	66	66
69											56	57	75	75			67	65	70	71	74	73
70											57	59					74	74	76	76		
71											60	60										

Note: SG indicates "single group," and EG, "equivalent groups" design

TABLE A-2

SAMPLE SIZES FOR EQUATING CAT TO PP ASVAB

	Equating design		
	Equivalent groups		
Subtest	CAT1	PP1	Single group
GS	1,130	1,189	2,319
AR	1,772	1,909	3,681
WK	1,937	2,104	4,041
PC	1,937	2,104	4,041
NO	1,373	1,303	2,676
CS	1,013	1,012	2,025
AS	1,479	1,468	2,947
MK	1,000	1,049	2,049
MC	1,546	1,587	3,133
EI	1,448	1,380	2,828
SI	1,470	1,459	2,929

END

Feb.

1988

DTIC